Journal of Multivariate Analysis 101 (2010) 949–963

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Boundary kernels for adaptive density estimators on regions with irregular boundaries

Jonathan C. Marshall, Martin L. Hazelton *

Institute of Fundamental Sciences – Statistics, Massey University, Palmerston North, New Zealand

ARTICLE INFO

Article history:
Received 5 November 2008
Available online 11 September 2009

AMS subject classifications:
62G07
62G20

Keywords:
Adaptive smoothing
Boundary bias
Edge effects
Kernel estimator
Variable bandwidth

ABSTRACT

In some applications of kernel density estimation the data may have a highly non-uniform distribution and be confined to a compact region. Standard fixed bandwidth density estimates can struggle to cope with the spatially variable smoothing requirements, and will be subject to excessive bias at the boundary of the region. While adaptive kernel estimators can address the first of these issues, the study of boundary kernel methods has been restricted to the fixed bandwidth context. We propose a new linear boundary kernel which reduces the asymptotic order of the bias of an adaptive density estimator at the boundary, and is simple to implement even on an irregular boundary. The properties of this adaptive boundary kernel are examined theoretically. In particular, we demonstrate that the asymptotic performance of the density estimator is maintained when the adaptive bandwidth is defined in terms of a pilot estimate rather than the true underlying density. We examine the performance for finite sample sizes numerically through analysis of simulated and real data sets.

1. Introduction

A common and important use of bivariate kernel density (or intensity) estimation is to describe the distribution of data across geographical regions. For example, Amatulli et al. [1] use density estimation to examine the occurrence of wildfires, while Benschop et al. [2] construct kernel estimates from data comprising the coordinates of cases of Salmonella infection on farms. Practical applications of this sort are characterized by two particular difficulties. First, standard kernel density estimators experience excessive bias at the boundaries of the region of interest. Second, the distribution of data is typically highly inhomogeneous (for instance, varying hugely between urban and rural regions), resulting in the type of smoothing problem for which fixed bandwidth kernel estimators are not well suited.

Methods exist for tackling these two difficulties separately. There is a significant body of literature on boundary bias correction for fixed bandwidth kernel density estimation, although the focus has been largely on the univariate case. See, for example, [3–6]. The options are more limited when it comes to correcting for edge effects in bivariate density estimation on geographical regions, since the boundaries are typically highly irregular. While this appears to discount the multivariate generalizations of a number of univariate methods (e.g. those based on beta and gamma kernels), it is nonetheless possible to craft suitable multivariate boundary kernels for fixed bandwidth density estimation using the techniques of Müller and Stadtmüller [7], Hazelton and Marshall [8], or Staniswalis et al. [9].

There is also a significant body of research on variable bandwidth kernel density estimators, which permit the degree of smoothing to adapt spatially according to the local requirements of the underlying density. See for example [10–12]. The most widely employed adaptive kernel density estimator is due to Abramson [13]. Abramson's method has excellent

* Corresponding author.
E-mail addresses: j.c.marshall@massey.ac.nz (J.C. Marshall), m.hazelton@massey.ac.nz (M.L. Hazelton).

0047-259X/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmva.2009.09.003

theoretical properties, and has also performed well in numerical experiments to investigate finite sample behaviour (e.g. [14]). It has also proven successful in dealing with the variable smoothing requirements of markedly inhomogeneous distributions in real data applications (e.g. [15,16]).

What is lacking in the current literature are methods which address the two difficulties simultaneously, i.e. techniques for correcting for boundary bias which can be employed with adaptive kernel estimators. In the univariate case there has been some progress, such as the transformation method of [17] and the variable location kernel density estimators of [18], however methods for the bivariate case are scarce. We address this important omission by developing a type of linear boundary kernel for use with Abramson's [13] adaptive density estimator for univariate and bivariate data. The leading term in the bias at the boundary (which is asymptotically non-zero for the uncorrected estimator) is quadratic in the (effective) bandwidth when this boundary kernel is employed.

The paper is structured as follows. In the next section we introduce preliminary material on adaptive kernel density estimation, and also lay the mathematical foundations for the asymptotic analysis of the properties of the density estimators at the boundary. The linear boundary kernels for univariate and bivariate cases are introduced in Section 3, and the asymptotic bias of the resulting density estimators given in a pair of theorems under the assumption that the adaptive local bandwidth is defined in a theoretically optimal (but impractical) manner. We then show that these asymptotic results continue to hold when the adaptive bandwidth is data-driven, being based upon a suitable pilot estimate of the underlying density. Finite sample performance is examined through numerical studies in Section 4 and we draw conclusions in Section 5. Proofs of theorems appear in a technical Appendix.

2. Preliminaries for density estimation at the boundary

2.1. Adaptive kernel density estimation

The kernel density estimator constructed from d-dimensional observations $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ can be written as

$$\bar{f}(\boldsymbol{x}) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{\boldsymbol{x} - \boldsymbol{X}_i}{h}\right). \tag{1}$$

Here K is the kernel function which we assume to be a spherically symmetric probability density, the support of which is the d-dimensional unit ball with $\int \|\boldsymbol{x}\|^2 K(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = d$. We will assume throughout the asymptotic analysis that K has at least two continuous derivatives. The parameter h is the bandwidth which controls the degree of smoothing of $\bar{f}$. To simplify notation henceforth, we define the scaled kernel by $K_h(\boldsymbol{x}) = h^{-d} K(\boldsymbol{x}/h)$. We will focus on the univariate and bivariate cases ($d = 1, 2$). For the latter, Eq. (1) can be generalized so as to incorporate a bandwidth matrix rather than a scalar h, permitting the amount of smoothing to change with direction (e.g. [19]). However, our particular interest is in applications where the data are Cartesian coordinates over a geographical region, for which isotropic smoothing kernels are natural.

Suppose that the observations are restricted to a (known) region $\mathcal{R}$, which is a compact connected set with a smooth boundary. For an interior point $\boldsymbol{x}$ the asymptotic mean and variance of $\bar{f}(\boldsymbol{x})$ are of order $O(h^2)$ and $O(n^{-1}h^{-d})$ respectively, so that it is optimal to set $h \propto n^{-1/4+d}$ when $\bar{f}(\boldsymbol{x})$ converges to $f(\boldsymbol{x})$ at rate $O(n^{-4/4+d})$. See for example [20] for details.

The estimator (1) uses a fixed bandwidth h at all locations. An alternative is to allow the degree of smoothing to adapt to local smoothing requirements. Arguably the most successful adaptive density estimator is due to [13], where the bandwidth varies with observation as follows:

$$\hat{f}(\boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^{n} h_i^{-d} K\left(\frac{\boldsymbol{x} - \boldsymbol{X}_i}{h_i}\right) \tag{2}$$

with

$$h_i = h(\boldsymbol{X}_i) = h_0 f(\boldsymbol{X}_i)^{-1/2}. \tag{3}$$

This choice of variable bandwidth makes intuitive sense in that it implies the use of larger bandwidths where there is likely to be less data and hence the need for a greater degree of smoothing. Theoretical motivation for (3) is provided by the improved asymptotic bias of the adaptive estimator. We have $\mathrm{Bias}[\hat{f}(\boldsymbol{x})] = O(h_0^4)$ for any point $\boldsymbol{x}$ interior to $\mathcal{R}$ while $\mathrm{Var}\{\hat{f}(\boldsymbol{x})\}$ remains of the same order, $O(n^{-1}h_0^{-1})$, as the variance for the fixed bandwidth estimator.

2.2. Foundations of asymptotic analysis at the boundary

Problems occur when $\boldsymbol{x}$ lies on the boundary $\partial\mathcal{R}$ of $\mathcal{R}$, since weight from kernels centred close to $\partial\mathcal{R}$ is lost over the boundary. A common approach to studying the properties of kernel density estimators close to the boundary is to consider estimation at a sequence of points that approach the $\partial\mathcal{R}$ as $n \to \infty$. In one dimension it is standard to fix $\alpha = x/h$ as $n \to \infty$ and $h = h_n \to 0$ (e.g. [3,4]). With this choice, the proportion of weight lost over the boundary remains constant for a kernel centred at x.

In higher dimensions a more sophisticated asymptotic analysis is required. We follow the approach of Müller and Stadtmüller [7] in which we consider the estimation point to be fixed while $\mathcal{R}$ changes with n. To describe this asymptotic

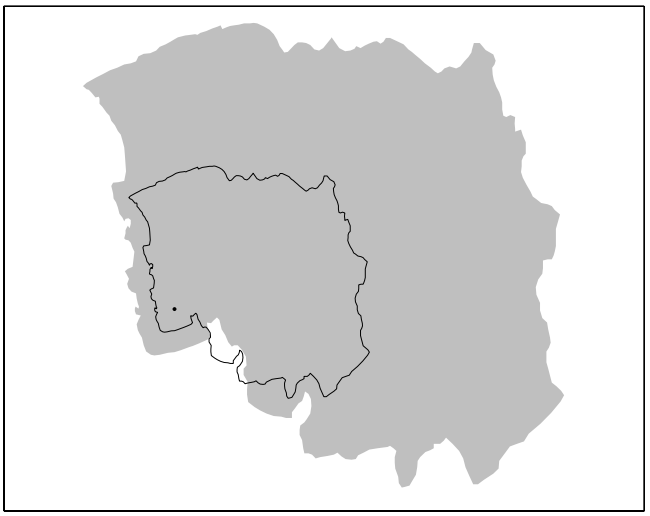


Fig. 1. $\mathcal{R}$ (filled) and $\mathcal{R}_n$ for a particular point $\boldsymbol{x}$ near the boundary.

device we introduce some additional notation. For a set A, let $A + \boldsymbol{x} = \{\boldsymbol{u} + \boldsymbol{x} : \boldsymbol{u} \in A\}$ and $\lambda A = \{\lambda \boldsymbol{u} : \boldsymbol{u} \in A\}$ for scalar λ. Also, we will use multi-index notation for vectors, so that if $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d)^\mathsf{T}$ then $\boldsymbol{x}^{\boldsymbol{\alpha}} = x_1^{\alpha_1} \cdots x_d^{\alpha_d}$, $|\boldsymbol{\alpha}| = \alpha_1 + \cdots + \alpha_d$, and

$$D^{\boldsymbol{\alpha}} f(\boldsymbol{x}) = (\partial^{\alpha_1}/\partial x_1^{\alpha_1}) \cdots (\partial^{\alpha_d}/\partial x_d^{\alpha_d}) f(\boldsymbol{x}).$$

Define $\mathcal{R}_n = \boldsymbol{x} + h(\mathcal{R} - \boldsymbol{x})$ where $h \to 0$ as $n \to \infty$, so that the distance of $\boldsymbol{x}$ from $\partial\mathcal{R}_n$ decreases in proportion to h (see Fig. 1). Let $\mathcal{T}_{n,\boldsymbol{x}} = \boldsymbol{x} - \mathcal{R}_n$ denote a translated version of the scaled region. Then

$$\int_{\mathcal{T}_{n,\boldsymbol{x}}} \boldsymbol{u}^{\boldsymbol{\alpha}} K_h(\boldsymbol{u}) \mathrm{d}\boldsymbol{u} = h^{|\boldsymbol{\alpha}|} a_{\boldsymbol{\alpha}} \tag{4}$$

where the coefficient

$$a_{\boldsymbol{\alpha}} = a_{\boldsymbol{\alpha}}(\boldsymbol{x}) = \int_{\mathcal{T}_{1,\boldsymbol{x}}} \boldsymbol{z}^{\boldsymbol{\alpha}} K(\boldsymbol{z}) \mathrm{d}\boldsymbol{z} \tag{5}$$

is independent of n.

Consider now the following sequence of problems indexed by n: estimate f restricted to the region $\mathcal{R}_n$ using the estimator (2) where terms for which $\boldsymbol{x}_i \notin \mathcal{R}_n$ are omitted. We can derive valid asymptotic expansions based on this limiting process courtesy of (4), which stabilizes the requisite asymptotic coefficients. In order to relate an expansion to a particular finite sample problem on region $\mathcal{R}$, using a bandwidth $\check{h} = \check{h}(\boldsymbol{x})$ thought of as fixed with respect to n, we prescale the problem by setting $\mathcal{R}_1 = \check{h}^{-1}\mathcal{R}$ so that the incomplete moment coefficients can be expressed by

$$a_{\boldsymbol{\alpha}} = \int_{\boldsymbol{x}-\mathcal{R}} (\boldsymbol{z}/\check{h})^{\boldsymbol{\alpha}} K_{\check{h}}(\boldsymbol{z}) \mathrm{d}\boldsymbol{z}.$$

Based on this asymptotic scheme, for a point $\boldsymbol{x}$ which is within a distance $h(\boldsymbol{x}) = h_0 f(\boldsymbol{x})^{-1/2}$ of $\partial\mathcal{R}$ we have

$$\mathsf{E}[\hat{f}(\boldsymbol{x})] = a_{\boldsymbol{0}} f(\boldsymbol{x}) + O(h) \tag{6}$$

where $\boldsymbol{0}$ is the d-vector of zeros. For example, if $d = 1$ and $\mathcal{R} = [0, L]$ then $\mathsf{E}[\hat{f}(0)] = \frac{1}{2} f(0) + O(h)$, which matches the well known result for univariate fixed kernel estimation at the boundary (e.g.[20]). Eq. (6) demonstrates that the bias in $\hat{f}(\boldsymbol{x})$ is so large when estimating at the boundary that the density estimator is rendered inconsistent.

A straightforward approach to correcting for boundary bias for the adaptive kernel estimator is simply to renormalize, giving $\hat{f}^R(\boldsymbol{x}) = \hat{f}(\boldsymbol{x})/a_{\boldsymbol{0}}(\boldsymbol{x})$. However, while $\hat{f}^R(\boldsymbol{x})$ is then consistent, this bias of this estimator is $O(h)$ near the boundary, which is significantly more severe than when estimating in the interior of $\mathcal{R}$.

3. Theory for linear boundary kernels

3.1. The univariate case

We can improve on the renormalization technique by using specially crafted boundary kernels to counteract the increased bias near $\partial\mathcal{R}$. Gasser and Müller [3] proposed a straightforward and effective class of boundary kernels for

univariate fixed bandwidth estimators, based on a linear combination of the original kernel K and $L(u) = uK(u)$. Specifically, for estimation at the point x, the linear boundary kernel is given by

$$B(u) = B_x(u) = \frac{(a_2 - a_1 u)K(u)}{a_0 a_2 - a_1^2} \, 1_{\mathcal{T}_{1,x}}(u).$$

More recently, Müller and Stadtmüller [7] examined multivariate boundary kernels for fixed bandwidth kernel estimators, using variational problems to generate optimal functional forms.

We develop boundary kernels for adaptive kernel estimators using the linear combinations approach. This process is more complicated than for the fixed bandwidth case because of the dependence of h on $f(x)$. For the univariate case we can obtain $O(h^2)$ bias for $\hat{f}$ using the boundary kernel

$$B(u) = \frac{[(a_3^{(1)} + 4a_2) - (a_2^{(1)} + 3a_1)u]K(u)}{(a_3^{(1)} + 4a_2)a_0 - (a_2^{(1)} + 3a_1)a_1} \, 1_{T_{1,x}}(u)$$

where

$$a_{\alpha}^{(\gamma)} = \int_{\mathcal{T}_{1,x}} \boldsymbol{z}^{\alpha} D^{\gamma} K(\boldsymbol{z}) \mathrm{d}\boldsymbol{z}.$$

We note that the kernel B is identical to K when estimating away from the boundary (i.e. when the support of K is contained within $\mathcal{T}_{1,x}$). Hence the adaptive kernel estimator constructed using the kernel function B, which we denote henceforth by $\hat{f}_B$, is identical to $\hat{f}$ (constructed with kernel K) away from $\partial\mathcal{R}$.

The asymptotic bias of $\hat{f}_B$ is given in the following theorem.

Theorem 1. *Assume that*
(A1) $f(x) \geq \delta > 0$ *for all* $x \in \mathcal{R}$;
(A2) $f(x)$ *has two continuous derivatives on* $\mathcal{R}$;
(A3) $h_i = h_0 f(\boldsymbol{X}_i)^{-1/2}$.

Then

$$\mathsf{E}[\hat{f}_B(x)] = f(x) + h(x)^2 \beta(x) + o(h_0^2),$$

where $h(x) = h_0 f(x)^{-1/2}$ *and*

$$\beta(x) = \frac{[D_1 f(x)]^2}{8f(x)} \cdot \frac{(a_3^{(1)} + 4a_2)(a_4^{(2)} + 5a_3^{(1)} + 3a_2) - (a_2^{(1)} + 3a_1)(a_5^{(2)} + 7a_4^{(1)} + 8a_3)}{(a_3^{(1)} + 4a_2)a_0 - (a_2^{(1)} + 3a_1)a_1} + \frac{f(x) D_2 f(x)}{8f(x)} \cdot \frac{(a_3^{(1)} + 4a_2)(2a_3^{(1)} + 6a_2) - (a_2^{(1)} + 3a_1)(2a_4^{(1)} + 8a_3)}{(a_3^{(1)} + 4a_2)a_0 - (a_2^{(1)} + 3a_1)a_1}.$$

Also, $\mathsf{Var}[\hat{f}_B(x)] = O(n^{-1} h_0^{-1})$.

The proof is given in the Appendix.

We note the assumption (A1) is sufficient (but not necessary) to ensure that the target density is always non-zero on $\partial\mathcal{R}_n$, and hence to rule out the trivial case where $f(x) = 0$. Further, the assumption (A2) of two continuous derivatives is required only for the calculation of the $O(h_0^2)$ term, and that with just one continuous derivative we have that $\mathsf{E}[\hat{f}_B(x)] = f(x) + O(h_0^2)$. Assumption (A3) is essentially a reminder that (for now) we have defined the local bandwidths in terms of the unknown f, in a manner that is theoretically optimal but impractical.

3.2. The bivariate case

Linear boundary kernels can also be constructed for bivariate density estimators. This approach has been studied in the context of fixed bandwidth estimators by Hazelton and Marshall [8]. By combining the original kernel K with $L_1(\boldsymbol{u}) = u_1 K(\boldsymbol{u})$ and $L_2(\boldsymbol{u}) = u_2 K(\boldsymbol{u})$ we obtain

$$B(\boldsymbol{u}) = B_{\boldsymbol{x}}(\boldsymbol{u}) = \frac{b_0 K + b_1 L_1(\boldsymbol{u}) + b_2 L_2(\boldsymbol{u})}{b_0 a_{00} + b_1 a_{10} + b_2 a_{01}} \, 1_{T_{1,\boldsymbol{x}}}(\boldsymbol{u}), \tag{7}$$

where

$$b_0 = (a_{30}^{(10)} + a_{21}^{(01)} + 5a_{20})(a_{12}^{(10)} + a_{03}^{(01)} + 5a_{02}) - (a_{21}^{(10)} + a_{12}^{(01)} + 5a_{11})(a_{21}^{(10)} + a_{12}^{(01)} + 5a_{11}),$$
$$b_1 = (a_{11}^{(10)} + a_{02}^{(01)} + 4a_{01})(a_{21}^{(10)} + a_{12}^{(01)} + 5a_{11}) - (a_{20}^{(10)} + a_{11}^{(01)} + 4a_{10})(a_{12}^{(10)} + a_{03}^{(01)} + 5a_{02}),$$
$$b_2 = (a_{20}^{(10)} + a_{11}^{(01)} + 4a_{10})(a_{21}^{(10)} + a_{12}^{(01)} + 5a_{11}) - (a_{11}^{(10)} + a_{02}^{(01)} + 4a_{01})(a_{30}^{(10)} + a_{21}^{(01)} + 5a_{20}).$$

By working with a linear combination of three kernels we are able to annihilate the $O(h)$ extra bias term that is present in the bivariate case when compared with the univariate one. As for the univariate setting, B is identical to the basic kernel K away from the boundary.

Theorem 2. *For $\boldsymbol{x} \in R$ let $\hat{f}_B(\boldsymbol{x})$ denote the bivariate adaptive kernel estimator constructed using the kernel function B. Then, under bivariate versions of the assumptions in Theorem 1,*

$$\mathsf{E}[\hat{f}_B(\boldsymbol{x})] = f(\boldsymbol{x}) + h(\boldsymbol{x})^2\beta(\boldsymbol{x}) + o(h_0^2),$$

where $h(\boldsymbol{x}) = h_0 f(\boldsymbol{x})^{-1/2}$ and

$$\begin{aligned}
\beta(\boldsymbol{x}) = {} & \frac{[D^{10}f(\boldsymbol{x})]^2}{4f(\boldsymbol{x})}\left[\frac{b_0}{q}\left(a_{40}^{(20)} + 2a_{31}^{(11)} + a_{22}^{(02)} + 7a_{30}^{(10)} + 7a_{21}^{(01)} + 8a_{20}\right)\right. \\
& + \frac{b_1}{q}\left(a_{50}^{(20)} + 2a_{41}^{(11)} + a_{32}^{(02)} + 9a_{40}^{(10)} + 9a_{31}^{(01)} + 15a_{30}\right) \\
& \left. + \frac{b_2}{q}\left(a_{41}^{(20)} + 2a_{32}^{(12)} + a_{23}^{(02)} + 9a_{31}^{(10)} + 9a_{22}^{(01)} + 15a_{21}\right)\right] \\
& + \frac{D^{10}f(\boldsymbol{x})D^{01}f(\boldsymbol{x})}{4f(\boldsymbol{x})}\left[\frac{b_0}{q}\left(2a_{31}^{(20)} + 4a_{22}^{(11)} + 2a_{13}^{(02)} + 14a_{21}^{(10)} + 14a_{12}^{(01)} + 16a_{11}\right)\right. \\
& + \frac{b_1}{q}\left(2a_{41}^{(20)} + 4a_{22}^{(11)} + 2a_{23}^{(02)} + 18a_{31}^{(10)} + 18a_{22}^{(01)} + 30a_{21}\right) \\
& \left. + \frac{b_2}{q}\left(2a_{32}^{(20)} + 4a_{23}^{(11)} + 2a_{14}^{(02)} + 18a_{22}^{(10)} + 18a_{13}^{(01)} + 30a_{12}\right)\right] \\
& + \frac{[D^{01}f(\boldsymbol{x})]^2}{4f(\boldsymbol{x})}\left[\frac{b_0}{q}\left(a_{22}^{(20)} + 2a_{13}^{(11)} + a_{04}^{(02)} + 7a_{12}^{(10)} + 7a_{03}^{(01)} + 8a_{02}\right)\right. \\
& + \frac{b_1}{q}\left(a_{32}^{(20)} + 2a_{23}^{(11)} + a_{14}^{(02)} + 9a_{22}^{(10)} + 9a_{13}^{(01)} + 15a_{12}\right) \\
& \left. + \frac{b_2}{q}\left(a_{23}^{(20)} + 2a_{14}^{(11)} + a_{05}^{(02)} + 9a_{13}^{(10)} + 9a_{04}^{(01)} + 15a_{03}\right)\right] \\
& + \frac{D^{20}f(\boldsymbol{x})}{4}\left[\frac{b_0}{q}\left(2a_{30}^{(10)} + 2a_{21}^{(01)} + 8a_{20}\right) + \frac{b_1}{q}\left(2a_{40}^{(10)} + 2a_{31}^{(01)} + 10a_{30}\right) + \frac{b_2}{q}\left(2a_{31}^{(10)} + 2a_{22}^{(01)} + 10a_{21}\right)\right] \\
& + \frac{D^{11}f(\boldsymbol{x})}{4}\left[\frac{b_0}{q}\left(4a_{21}^{(10)} + 4a_{12}^{(01)} + 16a_{11}\right) + \frac{b_1}{q}\left(4a_{31}^{(10)} + 4a_{22}^{(01)} + 20a_{21}\right) + \frac{b_2}{q}\left(4a_{22}^{(10)} + 4a_{13}^{(01)} + 20a_{12}\right)\right] \\
& + \frac{D^{02}f(\boldsymbol{x})}{4}\left[\frac{b_0}{q}\left(2a_{12}^{(10)} + 2a_{03}^{(01)} + 8a_{02}\right) + \frac{b_1}{q}\left(2a_{22}^{(10)} + 2a_{13}^{(01)} + 10a_{12}\right) + \frac{b_2}{q}\left(2a_{13}^{(10)} + 2a_{04}^{(01)} + 10a_{03}\right)\right],
\end{aligned}$$

where $q = b_0a_{00} + b_1a_{10} + b_2a_{01}$. Also, $\mathsf{Var}[\hat{f}_B(\boldsymbol{x})] = O(n^{-1}h_0^{-2})$.

The proof is given in the Appendix.

For both univariate and bivariate cases, the performance of $\hat{f}_B$ at the boundary remains worse than performance at interior points, where the bias is $O(h_0^4)$. Nonetheless, $\hat{f}_B$ has better properties than $\hat{f}^R$ for estimation on $\partial\mathcal{R}$. The variance of $\hat{f}_B$ remains of the same asymptotic order regardless of whether the estimating lies on the boundary or in the interior of $\mathcal{R}$. However, one needs to consider that, as with all linear boundary kernels, there is a possibility of negativity in the boundary region of the estimate $\hat{f}_B$. If this is undesirable, setting $\hat{f}_B$ to some minimal value at such points and rescaling to ensure the estimate integrates to 1 is a reasonably effective solution. Similarly, though the asymptotic variance is of the same order as the uncorrected estimate, for smaller sample sizes, the corrected estimates will in general have a larger variance close to $\partial\mathcal{R}$. There is thus a trade-off between having an inconsistent estimate or a corrected estimate with possibly larger variance that must be taken into account with such boundary kernels.

3.3. Pilot estimation

Implementation of Abramson's [13] adaptive estimator requires a pilot estimator $\tilde{f}$ of the target density in order to compute practical versions of the local bandwidths: $h_i = h_0\tilde{f}(\boldsymbol{X}_i)^{-1/2}$. Replacement of the true density by $\tilde{f}$ can lead to technical difficulties in the analysis of the properties of the resulting adaptive estimator $\hat{f}$, at least in part because of the non-local effects of huge bandwidths arising when $\tilde{f}(\boldsymbol{X}_i)$ is very small. See [21]. However, these problems can be avoided when the target density has compact support if we assume that f is bounded away on $\mathcal{R}$ (i.e. we assume that A1 holds from Theorem 1).

When this is the case, it can be shown that replacement of f by a suitable pilot estimator when computing the local bandwidths has no effect on the leading terms in the bias of the density estimator at the boundary. We provide a more formal statement of this result in the following theorem.

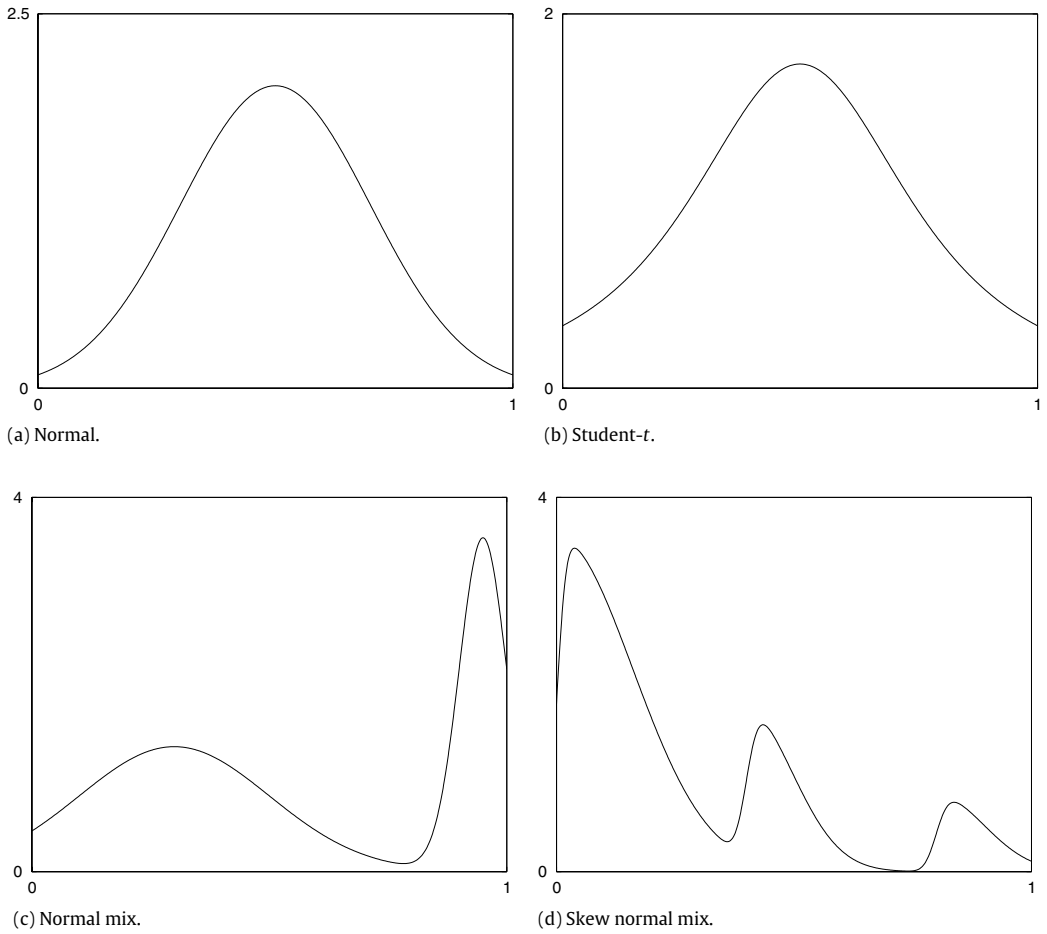


Fig. 2. Densities used for the one dimensional simulation study.

Theorem 3. *Let $\tilde{f}$ be a fixed bandwidth kernel estimator with boundary correction using bandwidth $\tilde{h} \sim n^{-1/d+4}$, and define the local bandwidth factors for the adaptive kernel estimator $\hat{f}_B$ by $h_i = h_0\tilde{f}(\boldsymbol{X}_i)^{-1/2}$. In addition assume that $\tilde{h}/h_0 \to 0$ as $n \to \infty$.*
Then the results of Theorems 1 and 2 continue to hold.

Our proof of this Theorem follows the general approach described by Hall and Marron [22], extended to the case where estimation is restricted to a finite region. The use of $\tilde{f}$ as the pilot estimator means we no longer have a symmetric kernel in the pilot estimate when close to the boundary, requiring a more delicate analysis of the asymptotic expansions. We also opt for a more direct approach, obtaining expressions for $\mathrm{E}[\hat{f}(\boldsymbol{x})]$ directly. The details are provided in the Appendix.

A suitable pilot would be a fixed bandwidth estimator (from (1)) implemented with linear boundary kernels, such as the those described in [8]. The assumption on $\tilde{h}/h_0$ is required only to obtain the same leading order term in the bias. This may be relaxed to $\tilde{h}/h_0 \to C$ where $C > 0$ is some finite constant and still give the same order of bias.

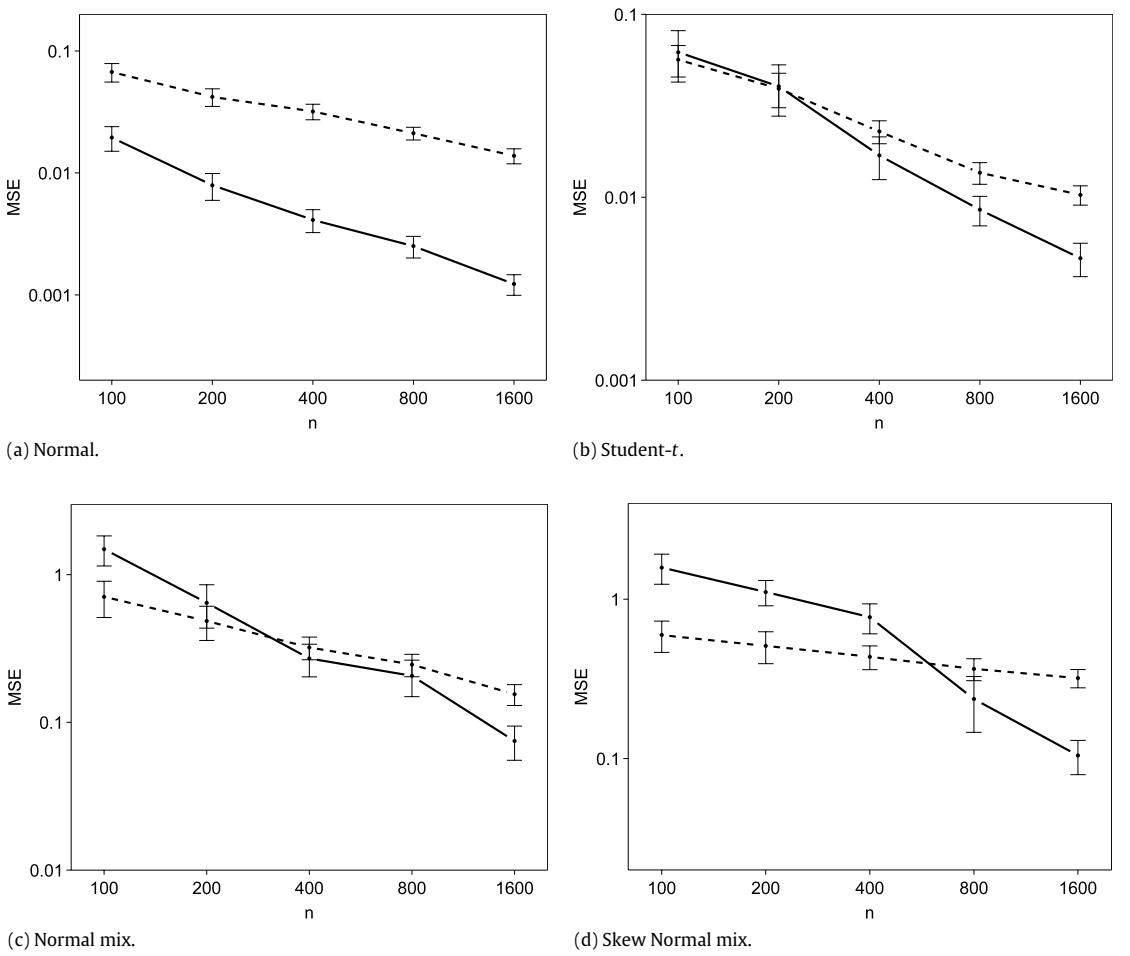
4. Numerical results

The numerical results derived in this section use the biweight kernel as the basic kernel K. Pilot estimation of f was done using a boundary corrected fixed bandwidth estimator $\tilde{f}$. For each data set considered, the global smoothing multiplier h_0 was selected using the cross-validation approach for adaptive estimation described in [23].

4.1. Simulation study: Univariate case

We assess the estimates $\hat{f}^R$ and $\hat{f}_B$ on simulated data sets from various distributions and compare their effectiveness in terms of mean square error from the true density at the boundary.

For the one dimensional case, consider the following distributions defined on the interval [0, 1], whose densities are shown in Fig. 2.



(a) Normal. (b) Student-t.

(c) Normal mix. (d) Skew Normal mix.

Fig. 3. Mean square error for $\hat{f}^R$ (dashed) and $\hat{f}_B$ (solid) evaluated on the boundary versus sample size. Error bars are 95% confidence intervals for the sample mean.

1. A normal distribution, $N(0.5, 0.2)$.
2. A student-t distribution, $0.25t(2) + 0.5$.
3. The normal mixture $0.6N(0.3, 0.2) + 0.4N(0.95, 0.05)$.
4. The skew normal mixture $0.7SN(0, 0.16, 10) + 0.2SN(0.4, 0.1, 5) + 0.1SN(0.8, 0.1, 5)$.

For each density we took samples of size $n = 100, 200, 400, 800$, and 1600, producing the estimates $\hat{f}^R$, and $\hat{f}_B$. The error in the each estimate was evaluated by averaging the square error of the estimates at $x = 0$ and $x = 1$. Sampling was repeated 100 times, and the results are presented in Fig. 3, where we give means and associated 95% confidence intervals.

The numerical results reflect the theory from Section 3. For each test density the linear boundary kernel approach provides better performance for sample sizes greater than 400. In some cases the improvement can be very large. Using $\hat{f}_B$ leads to a reduction in mean squared error at the boundary by an order of magnitude in comparison to $\hat{f}^R$ for $n \geq 1000$ for the normal test density, since the latter estimator overcorrects resulting in significant positive bias.

4.2. Simulation study: Bivariate case

For the two dimensional simulation study we consider four distributions restricted to the unit square $S = [0, 1] \times [0, 1]$. The densities used are

1. The bivariate normal $N([0.5, 0.5]^t; 0.04I_2)$.
2. The bimodal normal $0.5N([0.5, 0.13]^t; \sigma) + 0.5N([0.5, 0.87]^t; \sigma)$, where
$$\sigma = \begin{bmatrix} 0.02 & 0 \\ 0 & 0.01 \end{bmatrix}.$$
3. The normal mix $0.5N([0.2, 0.8]^t, 0.04I_2) + 0.15N([0.5, 0.05]^t, 0.0025I_2) + 0.35N([0.7, 0.6]^t, 0.01I_2)$.

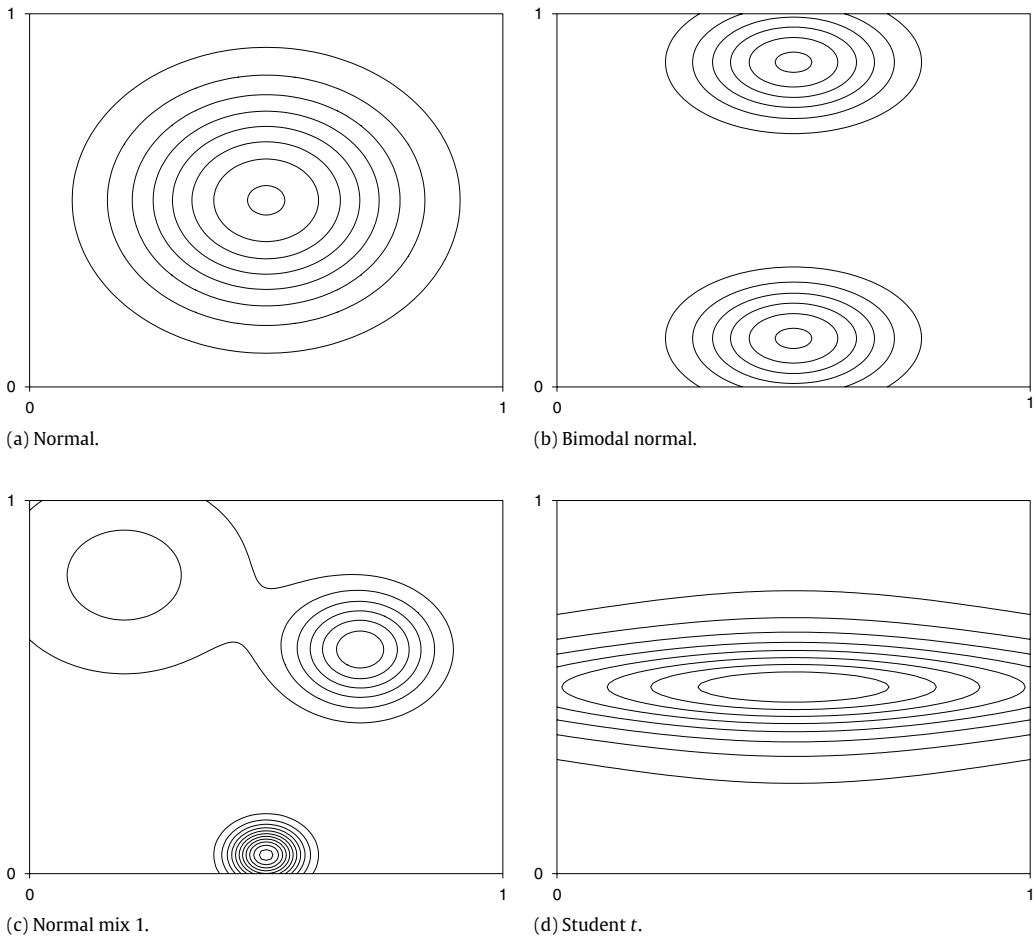


Fig. 4. Densities used for the two dimensional simulation study.

4. The Student t distribution $[0.5t(2) + 0.5] \times [0.1t(2) + 0.5]$.

These densities are plotted in Fig. 4.

Samples of size $n = 100$, 400, 1600, 6400, and 25 600, were taken from each distribution, and the estimates $\hat{f}^R$ and $\hat{f}_B$ were produced. The mean square error in each estimate at the boundary was computed by averaging the square error at each grid point on the boundary. Sampling was repeated 100 times, the results of which are given in Fig. 5.

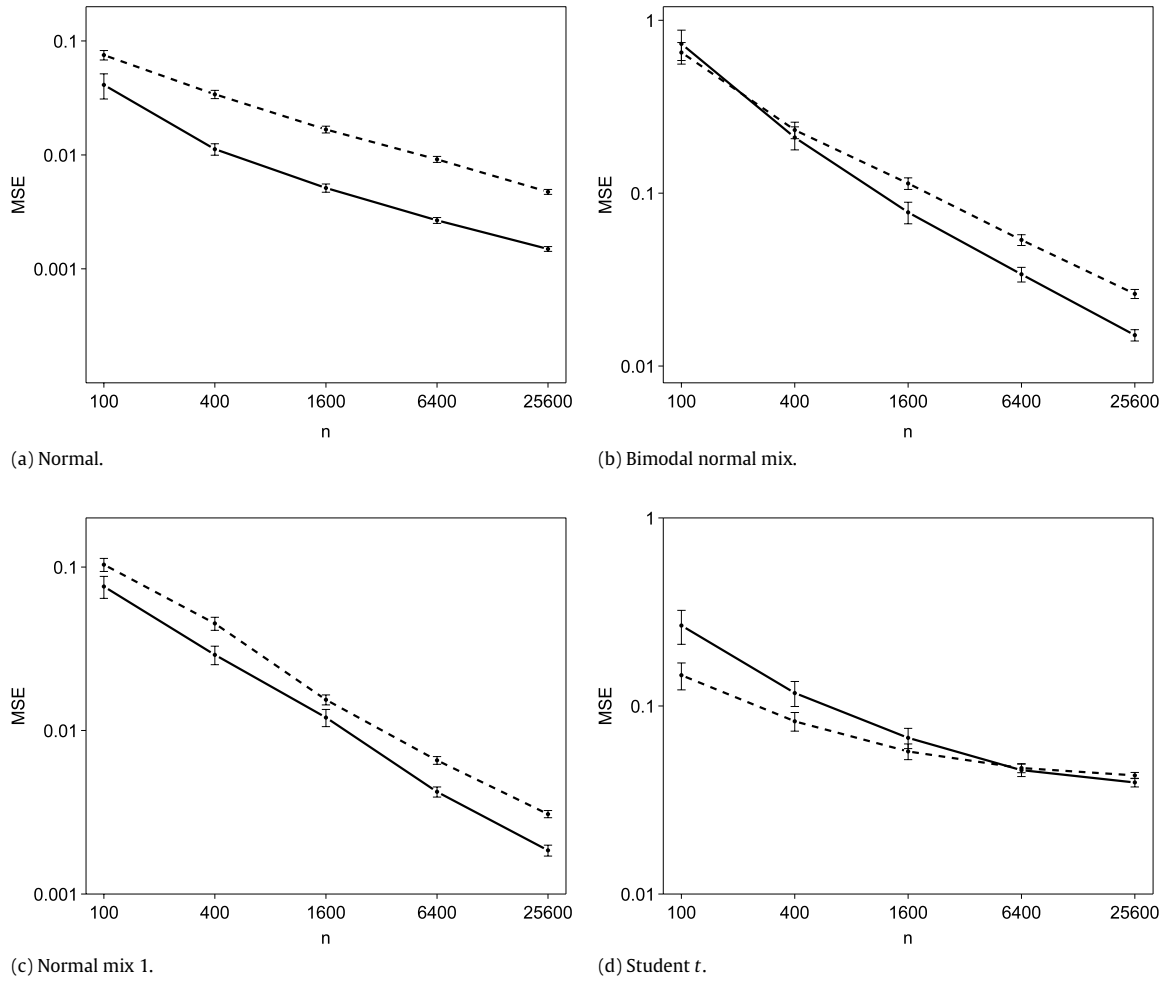
We once again see the theory reflected in the numerical results, with the linear boundary estimator $\hat{f}_B$ performing increasingly strongly in comparison with $\hat{f}^R$ as n increases. It is noticeable, however, that in some cases a large sample size may be required in order to realize the advantages of $\hat{f}_B$; see the bottom right-hand panel of Fig. 5, for example.

4.3. Data analyses

Our first example concerns the survival time in days following diagnosis of AIDS patients in Australia. The data consist of $n = 1761$ patients who were diagnosed with AIDS prior to 1 July 1991, with survival following diagnosis varying from 0 to 2252 days. See [24] for details.

We constructed an adaptive kernel density estimate using a biweight kernel (scaled to have unit variance) for the base kernel K, with a global bandwidth of $h_0 = 200$ days. In addition we constructed the boundary corrected estimates $\hat{f}^R$ and $\hat{f}_B$ using the same bandwidth. Fig. 6 displays these estimates as dotted, dashed and solid lines respectively. It is particularly clear from the right-hand panel (where we have focused on the value of the density near the boundary at $x = 0$) that linear boundary correction suggests a far steeper decline in the density over the first hundred days than does the rescaling approach.

For our second application, we have the geographical coordinates of $n = 62$ cases of childhood leukaemia and lymphoma collected between 1974 and 1986 in North Humberside, England. See [25] for details. The region concerned is approximately 70 km north–south by 60 km east–west, with an irregular polygonal boundary.



(a) Normal.

(b) Bimodal normal mix.

(c) Normal mix 1.

(d) Student t.

Fig. 5. Mean square error for $\hat{f}^R$ (dashed) and $\hat{f}_B$ (solid) on the boundary for increasing sample size. Error bars are 95% confidence intervals for the sample mean.

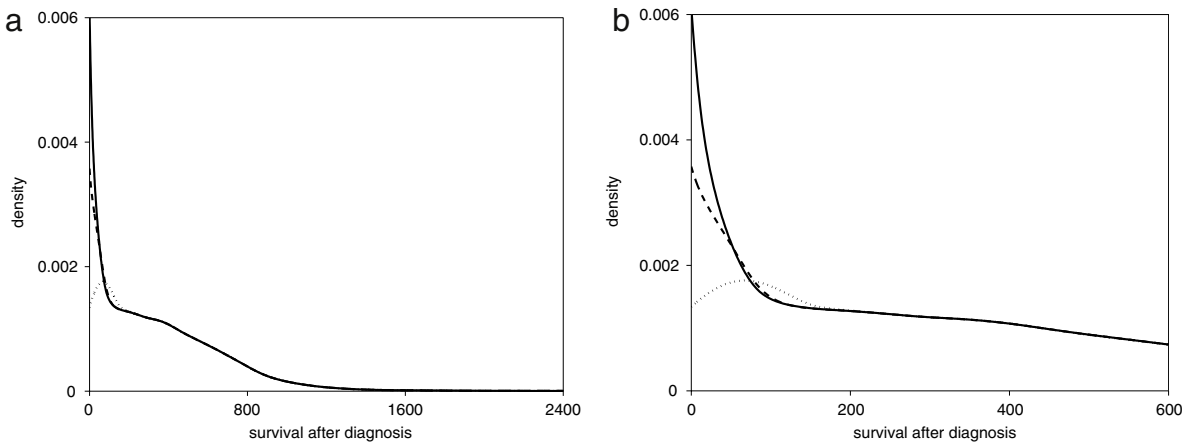


Fig. 6. Adaptive kernel density estimates $\hat{f}$ (dotted), $\hat{f}^R$ (dashed), and $\hat{f}_B$ (solid), from survival data on $n = 1761$ AIDS patients from Australia. The panel on the right focuses on the boundary at $x = 0$.

We constructed the adaptive kernel density estimates $\hat{f}$, $\hat{f}^R$ and $\hat{f}_B$ using a biweight kernel (scaled to have identity covariance matrix), and utilizing a global bandwidth of $h_0 = 3$ km. Fig. 7 gives perspective plots, with the left-hand panel presenting the standard adaptive estimate $\hat{f}$, the center panel presenting $\hat{f}^R$, and the right-hand panel showing the linear

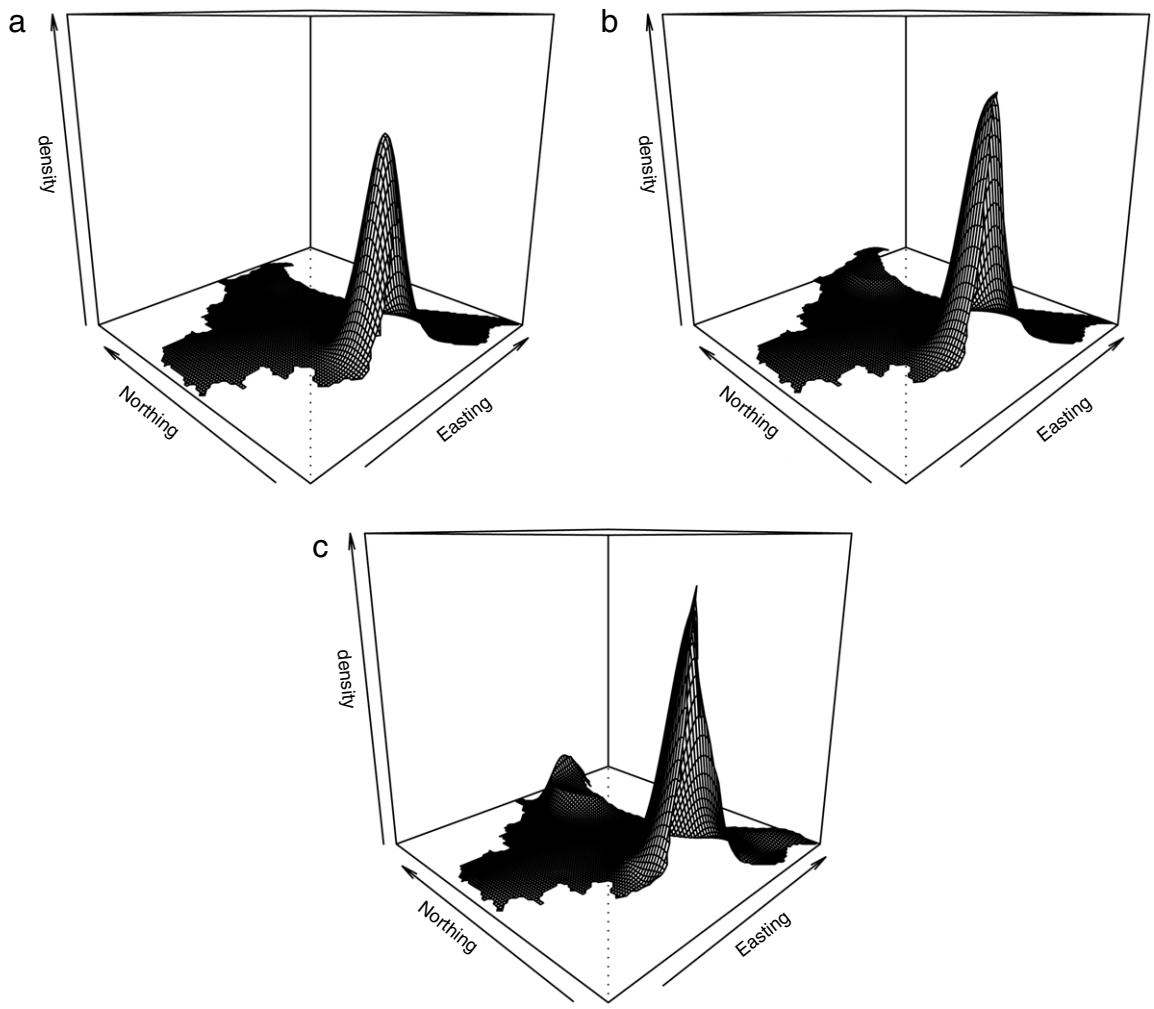


Fig. 7. Adaptive kernel density estimates from data on childhood leukaemia and lymphoma in North Humberside, England. The density on the top left is uncorrected, the density on the top right is corrected using a linear boundary kernel, and the density at the bottom is corrected using the renormalized kernel.

corrected estimate $\hat{f}_B$. The z-axis scale on each panel is the same. Of particular note is the peak in the north east, which is highlighted with the linear boundary corrected density more so than with $\hat{f}^R$, and is missed completely with the uncorrected density. It is clear that, even with a small data set, the linear boundary kernel is useful at picking up features of the data close to the boundary.

5. Conclusions

Adaptive kernel methods are attractive for estimating density or intensity functions relating to the distribution of populations of geographical regions. By introducing boundary kernels for these estimators and studying their theory, we have closed up a gap in the literature with significant practical consequences.

Our new adaptive boundary kernel is formed by a linear combination of a base kernel $K(\boldsymbol{x})$ with kernels of the form $L_i(\boldsymbol{x}) = x_i K(\boldsymbol{x})$. While this methodology is simple and effective, there are an unlimited number of alternative choices for L_i that will provide the same asymptotic order for the boundary bias of the adaptive kernel density estimator. The asymptotic coefficient, however, will be a complex amalgam of functionals of the density, the linear kernel, and the geometric properties of the boundary. In theory one could craft the boundary kernel for each application so as to minimize this coefficient. However, whether it would prove possible in practice to obtain a tangible benefit by doing so is an open question.

Finally, we note that while this paper has focused on boundary kernels for univariate and bivariate data, the ideas herein are readily extended to higher dimensional density estimation by including additional $x_i K(\boldsymbol{x})$ in the linear combination.

Acknowledgments

The authors thank two anonymous referees and an Associate Editor for their helpful comments.

Appendix. Proofs

Proof of Theorem 1. The expected value of (2) in the one dimensional case is given by

$$\mathsf{E}[\hat{f}(x)] = \int_{h_0^{-1}(x-\mathcal{R})} f(x+h_0z)^{3/2}K(zf(x+h_0z)^{1/2})\mathrm{d}z.$$

Replacing K with B and expanding the integrand as a Taylor series about $h_0 = 0$ yields

$$\begin{aligned}\mathsf{E}[\hat{f}_B(x)] &= f(x)\int_{\mathcal{T}_{1,x}} B(z)\mathrm{d}z + h(x)\frac{D^1f(x)}{2}\int_{\mathcal{T}_{1,x}} 3zB(z) + z^2D^1B(z)\mathrm{d}z \\ &\quad + h(x)^2\left\{\frac{[D^1f(x)]^2}{8f(x)}\int_{\mathcal{T}_{1,x}} z^4D^2B(z)\mathrm{d}z + \frac{5[D^1f(x)]^2 + 2f(x)D^2f(x)}{8f(x)}\int_{\mathcal{T}_{1,x}} z^3D^1B(z)\mathrm{d}z\right. \\ &\quad \left. + \frac{3[D^1f(x)]^2 + 6f(x)D^2f(x)}{8f(x)}\int_{\mathcal{T}_{1,x}} z^2B(z)\mathrm{d}z\right\} + o(h_0^2),\end{aligned}$$

where $h(x) = h_0f(x)^{-1/2}$. From the expression for B it is clear that

$$\int_{\mathcal{T}_{1,x}} B(z)\mathrm{d}z = 1,$$
$$\int_{\mathcal{T}_{1,x}} 3zB(z) + z^2B'(z)\mathrm{d}z = 0,$$

yielding the result in the theorem. The variance may be computed in the same way. We have

$$\begin{aligned}\mathsf{Var}\,\hat{f}_B(x) &= \frac{1}{nh_0}\int_{h_0^{-1}(x-\mathcal{R})} f(x+h_0z)^2B(zf(x+h_0z)^{1/2})\mathrm{d}z - \frac{1}{n}\left[\mathsf{E}\hat{f}_B(x)\right]^2 \\ &= \frac{f(x)}{nh(x)}\left\{\int_{\mathcal{T}_{1,x}} B(z)^2\mathrm{d}z + O(h_0)\right\} - \frac{1}{n}\{f(x) + O(h_0^2)\} \\ &= \frac{f(x)}{nh(x)}\int_{\mathcal{T}_{1,x}} B(z)^2\mathrm{d}z + O(n^{-1}),\end{aligned}$$

which is of the required order. □

Proof of Theorem 2. We follow the same technique used to prove Theorem 1. The expected value of (2) in two dimensions is given by

$$\mathsf{E}[\hat{f}_B(\boldsymbol{x})] = \int_{h_0^{-1}(\boldsymbol{x}-\mathcal{R})} f(\boldsymbol{x}+h_0\boldsymbol{z})^2B(\boldsymbol{z}f(\boldsymbol{x}+h_0\boldsymbol{z})^{1/2})\mathrm{d}\boldsymbol{z}.$$

Expanding the integrand as a Taylor series about $h_0 = 0$ yields

$$\begin{aligned}\mathsf{E}[\hat{f}_B(\boldsymbol{x})] &= f(\boldsymbol{x})c_{00} + h(\boldsymbol{x})\left[\frac{D^{10}f(\boldsymbol{x})}{2}\left(c_{20}^{(10)} + c_{11}^{(01)} + 4c_{10}\right) + \frac{D^{01}f(\boldsymbol{x})}{2}\left(c_{11}^{(10)} + c_{02}^{(01)} + 4c_{01}\right)\right] \\ &\quad + h(\boldsymbol{x})^2\left[\frac{[D^{10}f(\boldsymbol{x})]^2}{4f}\left(c_{40}^{(20)} + 2c_{31}^{(11)} + c_{22}^{(02)} + 7c_{30}^{(10)} + 7c_{21}^{(01)} + 8c_{20}\right)\right. \\ &\quad + \frac{D^{10}f(\boldsymbol{x})D^{01}f(\boldsymbol{x})}{4f(\boldsymbol{x})}\left(2c_{31}^{(20)} + 4c_{22}^{(11)} + 2c_{31}^{(02)} + 14c_{21}^{(10)} + 14c_{12}^{(01)} + 16c_{11}\right) \\ &\quad + \frac{[D^{01}f(\boldsymbol{x})]^2}{4f(\boldsymbol{x})}\left(c_{22}^{(20)} + 2c_{13}^{(11)} + c_{04}^{(02)} + 7c_{12}^{(10)} + 7c_{03}^{(01)} + 8c_{02}\right) \\ &\quad + \frac{D^{20}f(\boldsymbol{x})}{4}\left(2c_{30}^{(10)} + 2c_{21}^{(01)} + 8c_{20}\right) + \frac{D^{11}f(\boldsymbol{x})}{4}\left(4c_{21}^{(10)} + 4c_{12}^{(01)} + 16c_{11}\right) \\ &\quad \left. + \frac{D^{02}f(\boldsymbol{x})}{4}\left(2c_{12}^{(10)} + 2c_{03}^{(01)} + 8c_{02}\right)\right] + o(h_0^2),\end{aligned} \tag{8}$$

where $h(\boldsymbol{x}) = h_0 f(\boldsymbol{x})^{-1/2}$ and $c_{\boldsymbol{\alpha}}^{(\boldsymbol{\gamma})}$ are analogous to the $a_{\boldsymbol{\alpha}}^{(\boldsymbol{\gamma})}$ with $K(\boldsymbol{u})$ replaced with $B(\boldsymbol{u})$. Using (7) it is clear that

$$c_{ij}^{(kl)} = \frac{1}{q}\left[b_0 a_{ij}^{(kl)} + b_1 a_{(i+1)j}^{(kl)} + kb_1 a_{ij}^{((k-1)l)} + b_2 a_{i(j+1)}^{(kl)} + lb_2 a_{ij}^{(k(l-1))}\right],$$

with the convention that $a_{\boldsymbol{\alpha}}^{(00)} = a_{\boldsymbol{\alpha}}$. Using the expressions for b_i gives

$$\begin{aligned} &c_{00} = 1, \\ &c_{20}^{(10)} + c_{11}^{(01)} + 4c_{10} = 0, \\ &c_{11}^{(10)} + c_{02}^{(01)} + 4c_{01} = 0, \end{aligned}$$

and substitution into (8) yields the bias term in the theorem. The variance result is shown similarly. □

Proof of Theorem 3. Let $\check{f}_B(\boldsymbol{x})$ be the boundary corrected adaptive estimate using local bandwidths $h_i = h_0 \tilde{f}(\boldsymbol{X}_i)^{1/2}$, where $\tilde{f}(\boldsymbol{x})$ is a boundary corrected pilot estimator with kernel K and fixed bandwidth $\tilde{h}$. We show that

$$\mathsf{E}[\check{f}(\boldsymbol{x})] = \mathsf{E}[\hat{f}(\boldsymbol{x})] + O(\tilde{h}^2). \tag{9}$$

for points close to the boundary of $\mathcal{R}$, and

$$\mathsf{E}[\check{f}(\boldsymbol{x})] = \mathsf{E}[\hat{f}(\boldsymbol{x})] + O(\tilde{h}^2 h_0^2).$$

in the interior of $\mathcal{R}$.

Let $\delta(\boldsymbol{x}) = f(\boldsymbol{x}) - \tilde{f}(\boldsymbol{x})$ be the difference between the true density and pilot density. In what follows, the C_i's are assumed to be some constants. Then for d a positive integer,

$$\begin{aligned} \tilde{f}(\boldsymbol{y})^{d/2} &= \left[f(\boldsymbol{y})\left(1 - \frac{\delta(\boldsymbol{y})}{f(\boldsymbol{y})}\right)\right]^{d/2} \\ &= f(\boldsymbol{y})^{d/2}\left[1 - \frac{\mathrm{d}\delta(\boldsymbol{y})}{2f(\boldsymbol{y})} + \delta_1(\boldsymbol{y})\right] \end{aligned}$$

where

$$|\delta_1(\boldsymbol{y})| \le C_1 \delta(\boldsymbol{y})^2,$$

uniformly for $\boldsymbol{y} \in \mathcal{R}$. Thus

$$\begin{aligned} B\left(\frac{\boldsymbol{x}-\boldsymbol{y}}{h_0}\tilde{f}(\boldsymbol{y})^{1/2}\right) &= B\left(\frac{\boldsymbol{x}-\boldsymbol{y}}{h_0} f(\boldsymbol{y})\left[1 - \frac{\delta(\boldsymbol{y})}{2f(\boldsymbol{y})} + \delta_1(\boldsymbol{y})\right]\right) \\ &= B\left(\frac{\boldsymbol{x}-\boldsymbol{y}}{h_0} f(\boldsymbol{y})^{1/2}\right) - \frac{1}{2} B_1\left(\frac{\boldsymbol{x}-\boldsymbol{y}}{h_0} f(\boldsymbol{y})^{1/2}\right)\frac{\delta(\boldsymbol{y})}{f(\boldsymbol{y})} + \delta_2(\boldsymbol{x},\boldsymbol{y}) \end{aligned}$$

where $B_1(\boldsymbol{z}) = \sum_{i=1}^{d} z_i \frac{\partial}{\partial z_i} B(\boldsymbol{z})$ and

$$|\delta_2(\boldsymbol{x},\boldsymbol{y})| \le C_2 \delta(\boldsymbol{y})^2 I(\|\boldsymbol{x}-\boldsymbol{y}\| \le C_3 h_0),$$

uniformly for $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{R}$. Hence,

$$\begin{aligned} \tilde{f}(\boldsymbol{y})^{d/2} B\left(\frac{\boldsymbol{x}-\boldsymbol{y}}{h_0}\tilde{f}(\boldsymbol{y})^{/2}\right) &= f(\boldsymbol{y})^{d/2}\left[1 - \frac{\mathrm{d}\delta(\boldsymbol{y})}{2f(\boldsymbol{y})} + \delta_1(y)\right] \\ &\quad \times \left[B\left(\frac{\boldsymbol{x}-\boldsymbol{y}}{h_0} f(\boldsymbol{y})^{1/2}\right) - \frac{1}{2} B_1\left(\frac{\boldsymbol{x}-\boldsymbol{y}}{h_0} f(\boldsymbol{y})^{1/2}\right)\frac{\delta(\boldsymbol{y})}{f(\boldsymbol{y})} + \delta_2(\boldsymbol{x},\boldsymbol{y})\right] \\ &= f(\boldsymbol{y})^{d/2} B\left(\frac{\boldsymbol{x}-\boldsymbol{y}}{h_0} f(\boldsymbol{y})^{1/2}\right) - \frac{1}{2} f(\boldsymbol{y})^{d/2} L\left(\frac{\boldsymbol{x}-\boldsymbol{y}}{h_0} f(\boldsymbol{y})^{1/2}\right)\frac{\delta(\boldsymbol{y})}{f(\boldsymbol{y})} + \delta_3(\boldsymbol{x},\boldsymbol{y}) \end{aligned}$$

where $L(\boldsymbol{z}) = B(\boldsymbol{z}) + \mathrm{d}B_1(\boldsymbol{z})$ and

$$|\delta_3(\boldsymbol{x},\boldsymbol{y})| \le C_4 \delta(\boldsymbol{y})^2 I(\|\boldsymbol{x}-\boldsymbol{y}\| \le C_3 h_0),$$

uniformly for $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{R}$. We therefore have

$$\check{f}(\boldsymbol{x}) = \hat{f}(\boldsymbol{x}) - \frac{1}{2nh_0^d}\sum_{i=1}^{n} \delta(\boldsymbol{X}_i) f(\boldsymbol{X}_i)^{d/2-1} L\left(\frac{\boldsymbol{x}-\boldsymbol{X}_i}{h_0} f(\boldsymbol{X}_i)^{1/2}\right) + \delta_4(\boldsymbol{x})$$

where

$$\begin{aligned}|\delta_4(\boldsymbol{x})| &\le C_4\delta_5(\boldsymbol{x})\\ &= C_4\left\{\sup_{\boldsymbol{y}\in R}\delta(\boldsymbol{y})^2\right\}\frac{1}{nh_0^d}\sum_{i=1}^{n} I(\|\boldsymbol{x}-\boldsymbol{X}_j\|\le C_3h_0).\end{aligned}$$

We start by computing the expected value of the first term. Let

$$\begin{aligned}\epsilon(\boldsymbol{x}) &= \frac{1}{2nh_0^d}\sum_{i=1}^{n}\delta(\boldsymbol{X}_i)f(\boldsymbol{X}_i)^{d/2-1}L\left(\frac{\boldsymbol{x}-\boldsymbol{X}_i}{h_0}f(\boldsymbol{X}_i)^{1/2}\right)\\ &= \frac{1}{2n^2h_0^d}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[f(\boldsymbol{X}_i)^{d/2}-f(\boldsymbol{X}_i)^{d/2-1}\frac{1}{\tilde{h}^d}K\left(\frac{\boldsymbol{X}_i-\boldsymbol{X}_j}{\tilde{h}}\right)\right]L\left(\frac{\boldsymbol{x}-\boldsymbol{X}_i}{h_0}f(\boldsymbol{X}_i)^{1/2}\right)\\ &= \frac{1}{2n^2h_0^d}\sum_{i=1}^{n}\left[f(\boldsymbol{X}_i)^{d/2}-f(\boldsymbol{X}_i)^{d/2-1}\frac{K(0)}{\tilde{h}^d}\right]L\left(\frac{\boldsymbol{x}-\boldsymbol{X}_i}{h_0}f(\boldsymbol{X}_i)^{1/2}\right)\\ &\quad+\frac{1}{2n^2h_0^d}\sum_{i=1}^{n}\sum_{j\ne i}\left[f(\boldsymbol{X}_i)^{d/2}-f(\boldsymbol{X}_i)^{d/2-1}\frac{1}{\tilde{h}^d}K\left(\frac{\boldsymbol{X}_i-\boldsymbol{X}_j}{\tilde{h}}\right)\right]L\left(\frac{\boldsymbol{x}-\boldsymbol{X}_i}{h_0}f(\boldsymbol{X}_i)^{1/2}\right)\\ &= \frac{1}{2n}\epsilon_1(\boldsymbol{x})+\frac{1}{2}\epsilon_2(\boldsymbol{x}).\end{aligned}$$

Then

$$\begin{aligned}\mathsf{E}[\epsilon_1(\boldsymbol{x})] &= \int_{h_0^{-d}(\boldsymbol{x}-\mathcal{R})}\left[f(\boldsymbol{x}+h_0\boldsymbol{t})^{d/2+1}-\frac{K(0)}{\tilde{h}^d}f(\boldsymbol{x}+h_0\boldsymbol{t})^{d/2}\right]L(-\boldsymbol{t}f(\boldsymbol{x}+h_0\boldsymbol{t})^{1/2})\mathrm{d}\boldsymbol{t}\\ &= \left[f(\boldsymbol{x})-\frac{K(0)}{\tilde{h}^d}\right]\int_{\mathcal{T}_{1,\boldsymbol{x}}}L(-\boldsymbol{s})\mathrm{d}\boldsymbol{s}+\frac{h_0}{f(\boldsymbol{x})^{1/2}}\frac{f'(\boldsymbol{x})}{2}\int_{\mathcal{T}_{1,\boldsymbol{x}}}(d+2)\boldsymbol{s}L(-\boldsymbol{s})-\boldsymbol{s}^2L'(-\boldsymbol{s})\mathrm{d}\boldsymbol{s}\\ &\quad+\frac{h_0}{f(\boldsymbol{x})^{1/2}}\frac{K(0)f'(\boldsymbol{x})}{2\tilde{h}^df(\boldsymbol{x})}\int_{\mathcal{T}_{1,\boldsymbol{x}}}d\boldsymbol{s}L(-\boldsymbol{s})-\boldsymbol{s}^2L'(-\boldsymbol{s})\mathrm{d}\boldsymbol{s}+O(h_0^2),\end{aligned}$$

and

$$\begin{aligned}\mathsf{E}[\epsilon_2(\boldsymbol{x})] &= \frac{n-1}{n}\int_{h_0^{-d}(\boldsymbol{x}-\mathcal{R})}f(\boldsymbol{x}+h_0\boldsymbol{t})^{d/2}L(-\boldsymbol{t}f(\boldsymbol{x}+h_0\boldsymbol{t})^{1/2})\\ &\quad\times\left[f(\boldsymbol{x}+h_0\boldsymbol{t})-\int_{\tilde{h}^{-d}(\boldsymbol{x}+h_0\boldsymbol{t}-\mathcal{R})}K(-\boldsymbol{s})f(\boldsymbol{x}+h_0\boldsymbol{t}+\tilde{h}\boldsymbol{s})\mathrm{d}\boldsymbol{s}\right]\mathrm{d}\boldsymbol{t}\\ &= \frac{n-1}{n}\int_{h_0^{-d}(\boldsymbol{x}-\mathcal{R})}f(\boldsymbol{x}+h_0\boldsymbol{t})^{d/2}L(-\boldsymbol{t}f(\boldsymbol{x}+h_0\boldsymbol{t})^{1/2})\left[f(\boldsymbol{x}+h_0\boldsymbol{t})-\mathsf{E}[\tilde{f}(\boldsymbol{x}+h_0\boldsymbol{t},\tilde{h})]\right]\mathrm{d}\boldsymbol{t}\\ &= \frac{n-1}{n}\int_{h_0^{-d}(\boldsymbol{x}-\mathcal{R})}f(\boldsymbol{x}+h_0\boldsymbol{t})^{d/2}L(-\boldsymbol{t}f(\boldsymbol{x}+h_0\boldsymbol{t})^{1/2})O(\tilde{h}^2)\mathrm{d}\boldsymbol{t}\\ &= \frac{n-1}{n}O(\tilde{h}^2)\left[\int_{\mathcal{T}_{1,\boldsymbol{x}}}L(-\boldsymbol{s})\mathrm{d}\boldsymbol{s}+\frac{h_0}{f(\boldsymbol{x})^{1/2}}\frac{f'(\boldsymbol{x})}{2f(\boldsymbol{x})}\int_{\mathcal{T}_{1,\boldsymbol{x}}}d\boldsymbol{s}L(-\boldsymbol{s})-\boldsymbol{s}^2L'(-\boldsymbol{s})\mathrm{d}\boldsymbol{s}+O(h_0^2)\right].\end{aligned}$$

The integrals are at worst $O(1)$ on the boundary, and are zero in the interior. Thus

$$\mathsf{E}[\epsilon(\boldsymbol{x})]\sim O(\tilde{h}^2)$$

on the boundary, and

$$\mathsf{E}[\epsilon(\boldsymbol{x})]\sim O(\tilde{h}^2h_0^2)$$

in the interior.

It thus suffices to show that we may bound $\delta_4(\boldsymbol{x})$ by something smaller than $O(\tilde{h}^2h_0^2)$. Now

$$\delta_5(\boldsymbol{x})=\sup_{\boldsymbol{y}\in\mathcal{R}}\frac{1}{n^3}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\left[f(\boldsymbol{y})-\frac{1}{\tilde{h}^d}K\left(\frac{\boldsymbol{y}-\boldsymbol{X}_i}{\tilde{h}}\right)\right]\cdot\left[f(\boldsymbol{y})-\frac{1}{\tilde{h}^d}K\left(\frac{\boldsymbol{y}-\boldsymbol{X}_j}{\tilde{h}}\right)\right]\frac{1}{h_0^d}I(\|\boldsymbol{x}-\boldsymbol{X}_k\|\le C_3h_0)$$

so that

$$\mathsf{E}[\delta_5(\boldsymbol{x})] = \sup_{\boldsymbol{y}\in\mathcal{R}} \frac{1}{n^3}\{nI_1 + n(n-1)I_2 + 2n(n-1)I_3 + n(n-1)(n-2)I_4\},$$

where

$$I_1 = \int_{\mathcal{R}} \left[f(\boldsymbol{y}) - \frac{1}{\tilde{h}^d}K\left(\frac{\boldsymbol{y}-\boldsymbol{u}}{\tilde{h}}\right)\right]^2 \frac{1}{h_0^d} I(\|\boldsymbol{x}-\boldsymbol{u}\| \le C_3 h_0) f(\boldsymbol{u})\mathrm{d}\boldsymbol{u},$$

$$I_2 = \iint_{\mathcal{R}} \left[f(\boldsymbol{y}) - \frac{1}{\tilde{h}^d}K\left(\frac{\boldsymbol{y}-\boldsymbol{u}}{\tilde{h}}\right)\right]^2 \frac{1}{h_0^d} I(\|\boldsymbol{x}-\boldsymbol{v}\| \le C_3 h_0) f(\boldsymbol{u}) f(\boldsymbol{v})\mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v},$$

$$I_3 = \iint_{\mathcal{R}} \left[f(\boldsymbol{y}) - \frac{1}{\tilde{h}^d}K\left(\frac{\boldsymbol{y}-\boldsymbol{u}}{\tilde{h}}\right)\right]\left[f(\boldsymbol{y}) - \frac{1}{\tilde{h}^d}K\left(\frac{\boldsymbol{y}-\boldsymbol{v}}{\tilde{h}}\right)\right] \cdot \frac{1}{h_0^d} I(\|\boldsymbol{x}-\boldsymbol{u}\| \le C_3 h_0) f(\boldsymbol{u}) f(\boldsymbol{v})\mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v},$$

$$I_4 = \iiint_{\mathcal{R}} \left[f(\boldsymbol{y}) - \frac{1}{\tilde{h}^d}K\left(\frac{\boldsymbol{y}-\boldsymbol{u}}{\tilde{h}}\right)\right]\left[f(\boldsymbol{y}) - \frac{1}{\tilde{h}^d}K\left(\frac{\boldsymbol{y}-\boldsymbol{v}}{\tilde{h}}\right)\right] \cdot \frac{1}{h_0^d} I(\|\boldsymbol{x}-\boldsymbol{w}\| \le C_3 h_0) f(\boldsymbol{u}) f(\boldsymbol{v}) f(\boldsymbol{w})\mathrm{d}\boldsymbol{u}\mathrm{d}\boldsymbol{v}\mathrm{d}\boldsymbol{w},$$

Setting

$$\beta_i = \int_{\mathcal{R}} \frac{1}{\tilde{h}^{di}}K\left(\frac{\boldsymbol{y}-\boldsymbol{u}}{\tilde{h}}\right)^i \frac{1}{h_0^d} I(\|\boldsymbol{x}-\boldsymbol{u}\| \le C_3 h_0) f(\boldsymbol{u})\mathrm{d}\boldsymbol{u},$$

gives

$$I_1 = f(\boldsymbol{y})^2\beta_0 - 2f(\boldsymbol{y})\beta_1 + \beta_2,$$
$$I_2 = f(\boldsymbol{y})^2\beta_0 - 2f(\boldsymbol{y})\mathsf{E}[\tilde{f}(\boldsymbol{y})]\beta_0 + \mathsf{E}[\tilde{f}(\boldsymbol{y})^2]\beta_0,$$
$$I_3 = f(\boldsymbol{y})^2\beta_0 - f(\boldsymbol{y})\left[\beta_1 + \mathsf{E}[\tilde{f}(\boldsymbol{y})]\beta_0\right] + \mathsf{E}[\tilde{f}(\boldsymbol{y})]\beta_1,$$
$$I_4 = f(\boldsymbol{y})^2\beta_0 - 2f(\boldsymbol{y})\mathsf{E}[\tilde{f}(\boldsymbol{y})]\beta_0 + \mathsf{E}[\tilde{f}(\boldsymbol{y})]^2\beta_0,$$

so that

$$\mathsf{E}[\delta_5(\boldsymbol{x})] = \sup_{\boldsymbol{y}\in\mathcal{R}} \frac{1}{n^3}\left\{\left[n^3 f(\boldsymbol{y})^2 - 2(n^3-n^2)f(\boldsymbol{y})\mathsf{E}[\tilde{f}(\boldsymbol{y})] + (n^3-2n^2+n)\mathsf{E}[\tilde{f}(\boldsymbol{y})]^2\right]\beta_0 + \left[-2n^2 f(\boldsymbol{y}) + 2(n^2-n)\mathsf{E}[\tilde{f}(\boldsymbol{y})]\right]\beta_1 + n\beta_2\right\}.$$

Given that

$$\mathsf{E}[\tilde{f}(\boldsymbol{y})] = f(\boldsymbol{y}) + O(\tilde{h}^2)$$
$$\mathsf{E}[\tilde{f}(\boldsymbol{y})^2] = \frac{1}{n\tilde{h}^d} f(\boldsymbol{y}) \int_{\mathcal{T}_{1,\boldsymbol{x}}} K(\boldsymbol{s})^2 \mathrm{d}\boldsymbol{s} + O(n^{-1})$$

we have

$$\mathsf{E}[\delta_5(\boldsymbol{x})] \sim \frac{1}{n^3}\left\{O(n^2)\beta_0 + O(n^2\tilde{h}^2)\beta_1 + n\beta_2\right\}.$$

It is clear that $\beta_0 \sim O(1)$, and $\beta_i \le \frac{1}{h_0^d}\mathsf{E}[\tilde{f}(\boldsymbol{y})^i]$ for $i \in \{1, 2\}$. Thus,

$$\mathsf{E}[\delta_4(\boldsymbol{x})] \sim O(n^{-1}),$$

as required. □

References

[1] G. Amatulli, F. Perez-Cabello, J. de la Riva, Mapping lightning/human-caused wildfires occurrence under ignition point location uncertainty, Ecological Model. 200 (2007) 321–333.
[2] J. Benschop, M.L. Hazelton, M.A. Stevenson, J. Dahl, R.S. Morris, N.P. French, Descriptive spatial epidemiology of subclinical salmonella infection in danish finisher pig herds: Application of a novel method of spatially adaptive smoothing, Veterinary Research 39:02.
[3] T. Gasser, H. Müller, Kernel estimation of regression functions, in: Smoothing Techniques for Curve Estimation (Proc. Workshop, Heidelberg, 1979), in: Lecture Notes in Math., vol. 757, Springer, Berlin, 1979, pp. 23–68.
[4] T. Gasser, H. Müller, V. Mammitzsch, Kernels for nonparametric curve estimation, J. Roy. Statist. Soc. Ser. B 47 (2) (1985) 238–252.
[5] S.X. Chen, Beta kernel estimators for density functions, Comput. Statist. Data Anal. 31 (2) (1999) 131–145.

[6] S.X. Chen, Probability density function estimation using gamma kernels, Ann. Inst. Statist. Math. 52 (3) (2000) 471–480.
[7] H.G. Müller, U. Stadtmüller, Multivariate boundary kernels and a continuous least squares principle, J. Roy. Statist. Soc. Ser. B 61 (2) (1999) 439–458.
[8] M.L. Hazelton, J.C. Marshall, Linear boundary kernels for bivariate density estimation, Statist. Probab. Lett. 79 (2009) 999–1003.
[9] J. Staniswalis, K. Messer, D. Finston, Kernel estimators for multivariate regression, J. Nonparametr. Stat. 3 (2) (1993) 103–121.
[10] S.R. Sain, D.W. Scott, On locally adaptive density estimation, J. Amer. Statist. Assoc. 91 (436) (1996) 1525–1534.
[11] M.L. Hazelton, Variable kernel density estimation, Aust. N.Z. J. Stat. 45 (3) (2003) 271–284.
[12] S.R. Sain, Multivariate locally adaptive density estimation, Comput. Statist. Data Anal. 39 (2) (2002) 165–186.
[13] I. Abramson, On bandwidth variation in kernel estimation – a square root law, Ann. Statist. 10 (1982) 1217–1223.
[14] S. Sorbye, F. Godtliebsen, Finite sample properties of an adaptive density estimator, J. Nonparametr. Stat. 14 (4) (2002) 383–398.
[15] D. Fadda, E. Slezak, A. Bijaoui, Density estimation with non-parametric methods, Astronomy Astrophys. Suppl. Ser. 127 (2) (1998) 335–352.
[16] R. Ezcurra, Is there cross-country convergence in carbon dioxide emissions? Energy Policy 35 (2007) 1363–1372.
[17] R. Karunamuni, T. Alberts, A locally adaptive transformation method of boundary correction in kernel density estimation, J. Statist. Plann. Inference 136 (9) (2006) 2936–2960.
[18] B.U. Park, S.O. Jeong, M.C. Jones, K.H. Kang, Adaptive variable location kernel density estimators with good performance at boundaries, J. Nonparametr. Stat. 15 (15) (2003) 61–75.
[19] T. Duong, M.L. Hazelton, Plug-in bandwidth matrices for bivariate kernel density estimation, J. Nonparametr. Stat. 15 (1) (2003) 17–30.
[20] M. Wand, M. Jones, Kernel Smoothing, Chapman & Hall, London, 1995.
[21] P. Hall, T.C. Hu, J.S. Marron, Improved variable window kernel estimates of probability densities, Ann. Statist. 23 (1) (1995) 1–10.
[22] P. Hall, J.S. Marron, Variable window kernel estimates of probability densities, Probab. Theory Related Fields 80 (1988) 37–49.
[23] B. Silverman, Density Estimation for Statistics and Data Analysis, Chapman & Hall, New York, 1986.
[24] W.N. Venables, B.D. Ripley, Modern Ann. Statist. with S, 4th ed., Springer, 2002.
[25] J. Cuzick, R. Edwards, Spatial clustering for inhomogeneous populations, J. Roy. Statist. Soc. Ser. B 52 (1990) 73–104.